

Министерство науки и высшего образования Российской Федерации
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ (НИ ТГУ)

Механико-математический факультет

УТВЕРЖДЕНО:
Декан ММФ ТГУ
Л.В.Гензе

Оценочные материалы по дисциплине

Статистический анализ данных

по направлению подготовки

01.03.01 Математика

02.03.01 Математика и компьютерные науки

Направленность (профиль) подготовки

Основы научно-исследовательской деятельности в области математики
Основы научно-исследовательской деятельности в области математики
и компьютерных наук

Форма обучения

Очная

Квалификация

Бакалавр

Год приема

2023

СОГЛАСОВАНО:
Руководитель ОП
Л.В.Гензе

Председатель УМК
Е.А.Тарасов

1. Компетенции и индикаторы их достижения, проверяемые данными оценочными материалами

Целью освоения дисциплины является формирование следующих компетенций:

ОПК-4 Способен проводить под научным руководством исследование на основе существующих методов в конкретной области профессиональной деятельности.

ОПК-8 Способен использовать в педагогической деятельности научные знания в сфере математики, механики, компьютерных наук и информатики.

ПК-1 Способен проводить научно-исследовательские разработки по отдельным разделам выбранной темы.

Результатами освоения дисциплины являются следующие индикаторы достижения компетенций:

ИОПК 4.1 Проводит поиск и обработку научной и научно-технической информации, необходимой для решения исследовательских задач

ИОПК 4.2 Оценивает полученные результаты и формулирует выводы по итогам проведенных исследований

ИОПК 8.1 Демонстрирует способность подготовить конспект или план занятия по теме из области математики, механики, компьютерных наук или информатики.

ИОПК 8.2 Выбирает подходящие источники информации для подготовки конспекта или плана занятия по выбранной теме.

ИПК 1.1 Проводит работы по обработке и анализу научно-технической информации и результатов исследований

ИПК 1.2 Подготавливает планы и программы проведения отдельных этапов научно-исследовательской работы

ИПК 1.3 Проводит отдельные этапы научно-исследовательской работы

2. Оценочные материалы текущего контроля и критерии оценивания

Элементы текущего контроля:

- реферат;
- практические задачи;
- опросы на лекции.

Темы рефератов (ИОПК 8.1, ИОПК 8.2, ИПК 1.1):

- 1) Модель парной линейной регрессии. Метод наименьших квадратов.
- 2) Модель Лассо для зависимых данных. Проверка качества модели.
- 3) Пробит-модель, проверка ее качества.
- 4) Бутстрап-методы.

Критерии оценивания рефератов:

За рефераты выставляется оценка «зачтено» или «незачтено».

Оценка «зачтено» ставится, если реферат подготовлен в полном объеме и раскрывает полностью содержание заявленной темы.

Оценка «незачтено» выставляется в остальных случаях.

Примеры практических заданий (ИОПК 4.1, ИОПК 4.2, ИПК 1.2, ИПК 1.3):

Задача №1.

Себестоимость тонны угля Y в зависимости от среднемесячной производительности труда шахтера X для одной из шахт страны представлена в таблице

X	210	240	280	300	340	350	360	390	400
Y	20	13	12	13	11	11	10	11	10

Подобрать форму зависимости Y от X и оценить параметры полученной регрессионной модели.

Задача №2. Для векторов выборочных данных

Y	12.2	7.6	10.4	9.9	15.7	14
X_1	4795	6962	6571	4249	9540	3488
X_2	69	82	87	92	23	31
Y	12.7	10.5	15.1	10.6	15.2	17.2
X_1	4888	6237	2997	2990	1748	2128
X_2	55	81	65	98	100	69

Построить линейную модель множественной регрессии, проверить ее качество, оставить лишь значимые предикторы. Построить графическую иллюстрацию построенной модели и облако точек.

Задача №3.

Фирма 1 заключила договор с фирмой 2, предметом которого является оказание услуг связи, в том числе и доступ в Интернет. С целью увеличения интеллектуального потенциала фирмы руководство решило расширить круг своих сотрудников, обладающих правом бесплатного доступа. Выбор таких сотрудников решено было осуществлять с учетом ряда факторов, тем или иным образом характеризующих претендентов с точки зрения эффективного использования Интернет-ресурсов. В связи с этим возник вопрос: Кому из претендентов на бесплатный доступ предоставить такую возможность в первую очередь? Для того, чтобы получить обоснованный ответ, руководство фирмы поручило экономико-аналитическому отделу разработать модель, позволяющую по каждому претенденту рассчитать прогнозную оценку целесообразности предоставления ему права бесплатного доступа к ресурсам Интернета. В основу построения такой модели была положена идея применения бинарной переменной Y : $Y=1$ если сотрудник, обладающий правом бесплатного доступа к Интернет-ресурсам, по оценке экспертной группы, эффективно пользуется этим правом; $Y = 0$, в противном случае. Построение модели руководство фирмы предложило провести по пяти факторам: возраст (X_1 , лет), стаж профессиональной деятельности (X_2 , лет), заработная плата (X_3 , тыс. ден. ед.), число случаев поступления полезной для фирмы информации от сотрудника (X_4 , ед.), результат тестирования на предмет оценки навыков работы в Интернет (X_5 , в баллах). Значения этих показателей, а также значения бинарной переменной для 100 сотрудников фирмы представлены в таблице 1.1.

Таблица 1.1

№	Y	X_1	X_2	X_3	X_4	X_5
1	0	22	1	2,5	2	6
2	0	24	1	3	3	8
3	1	25	1	2,1	1	7
4	1	27	4	4,6	8	11
5	1	28	3	5,9	9	13
6	0	21	1	3,3	5	
7	1	22	1	3,2	9	15
8	0	29	5	6	1	13
9	1	26	4	2,7	8	11
10	0	27	6	3,8	0	3

По имеющимся данным построить модель логистической регрессии, оценить параметры модели, вычислить коэффициент детерминации, проверить значимость модели.

Критерии оценивания:

Результаты решения практических задач определяются «отлично», «хорошо», «удовлетворительно», «неудовлетворительно».

Оценка «отлично» выставляется, если в задаче применены необходимые статистические методы для ее решения, получено обоснованное верное решение, и студент умеет интерпретировать полученный результат с использованием соответствующей теории.

Оценка «хорошо» выставляется, если в задаче применены необходимые статистические методы для ее решения, получено обоснованное верное решение, но не все этапы решения объяснены теоретически, либо решение содержит несущественные ошибки;

Оценка «удовлетворительно» выставляется, если студент верно решил задачу, но не может обосновать решение теоретически.

Оценка «неудовлетворительно» выставляется, если задача решена неверно.

3. Оценочные материалы итогового контроля (промежуточной аттестации) и критерии оценивания

Критерии оценивания:

Зачет проводится в письменной форме с учетом работы студентов в семестре. Билет содержит один теоретический вопрос, проверяющий ИОПК 4.2, ИПК 1.3.

Пример билета.

- Процедура сглаживания данных методом скользящей медианы.

Результаты дифференцированного зачета определяются оценками «отлично», «хорошо», «удовлетворительно», «неудовлетворительно».

Студенту ставится оценка:

- 1) «удовлетворительно», если студент выполнил все задания на минимальную оценку «удовлетворительно» и ответил на теоретический вопрос в билете, понимая суть вопроса, но со значительными пробелами в выводах;
- 2) «хорошо», если все задачи выполнены в среднем на оценку «хорошо» и студент ответил на теоретический вопрос, но в ответе были мелкие неточности и не ответил на дополнительные вопросы, то есть ответ был неполным;
- 3) «отлично» ставится, если студент выполнил все практические задания на оценку «отлично» и ответил на теоретический вопрос достаточно развернуто и без ошибок.
- 4) «неудовлетворительно», если не выполнены критерии, описанные в пунктах 1) -3).

4. Оценочные материалы для проверки остаточных знаний (сформированности компетенций)

Оценочные материалы для проверки остаточных знаний могут быть использованы для формирования программы ГИА (программы государственного экзамена), а также экспертом Рособрнадзора при проведении проверки диагностической работы по оценке уровня форсированности компетенций обучающихся (при контрольно-надзорной проверке). Вопросы данного раздела показывают вклад дисциплины в образовательный результат образовательной программы. Объем заданий в данном разделе зависит как от

количества формируемых индикаторов достижения компетенций, так и от объема дисциплины по учебному плану.

Теоретические вопросы по курсу (ИОПК 4.1, ИОПК 4.2, ИПК 1.3):

1. Модель парной линейной регрессии. Оценивание параметров модели методом наименьших квадратов. Проверка качества модели.
2. Модель множественной регрессии. Оценивание параметров модели. Свойства полученных оценок. Проверка качества модели. Доверительные интервалы для параметров модели.
3. Мультиколлинеарность данных. Обнаружение мультиколлинеарности с помощью собственных чисел. Построение VIF.
4. Модель ридж-регрессии. Подбор параметра регуляризации.
5. Модель Лассо.
6. Робастные модели регрессии при наличии выбросов. Использование M-оценок.
7. Сглаживание данных методом скользящей медианы.
8. Логит-модель, оценивание ее параметров. Проверка качества модели.
9. Пробит-модель. Оценка параметров данной модели. Меры адекватности. ROC-кривая и AUC.
10. Байесовский подход. Априорные распределенные, сопряженные с генеральной совокупностью. САЗ – распределения. Переход от априорного распределения параметров к апостериорному с помощью формулы Байеса.
11. Байесовский прогноз отклика для модели линейной регрессии в предположении нормальности остатков.
12. Бутстрапирование. Бутстраповская корректировка смещения. Ресэмплинг в регрессии.

Ответ должен содержать теорию, изложенную в лекциях, в объеме основных понятий по полученному вопросу.

Информация о разработчиках

Губин Владимир Николаевич, кандидат физико-математических наук, ТГУ, доцент.